Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Understanding enzyme function evolution from a computational perspective

Jonathan D Tyzack[1], Nicholas Furnham[2], Ian Sillitoe[3], Christine M Orengo[3] and Janet M Thornton[1]

In this review, we will explore recent computational approaches to understand enzyme evolution from the perspective of protein structure, dynamics and promiscuity. We will present quantitative methods to measure the size of evolutionary steps within a structural domain, allowing the correlation between change in substrate and domain structure to be assessed, and giving insights into the evolvability of different domains in terms of the number, types and sizes of evolutionary steps observed. These approaches will help to understand the evolution of new catalytic and non-catalytic functionality in response to environmental demands, showing potential to guide *de novo*enzyme design and directed evolution experiments.

Addresses
[1] EMBL-EBI, Wellcome Genome Campus, CB10 1SD, United Kingdom
[2] London School of Hygiene and Tropical Medicine, Keppel Street, London, WC1E 7HT, United Kingdom
[3] Institute of Structural and Molecular Biology, University College London, Gower Street, London, WC1E 6BT, United Kingdom

Corresponding author: Tyzack, Jonathan D (tyzack@ebi.ac.uk)

Current Opinion in Structural Biology 2017, **47**:131–139

This review comes from a themed issue on **Catalysis and regulation**

Edited by **Christine Orengo** and **Janet Thornton**

For a complete overview see the Issue and the Editorial

Available online 8th September 2017

http://dx.doi.org/10.1016/j.sbi.2017.08.003

0959-440/© 2017 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Introduction

Enzymes are the product of billions of years of evolution giving us molecular machines that are critical for life. Across the vast space of protein structure, evolution has settled upon a limited number of structural folds that support an incredibly diverse chemistry acting on a multitude of substrates. Enzymes have evolved substrate and function specificity to improve the organism's overall fitness in response to environmental demands, but enzymes can have multiple substrates and functions, contradicting the traditional view of one enzyme one reaction and the high specificity implied in the lock and key and induced-fit paradigms. Furthermore, it is advantageous for organisms to be able to adapt quickly to a changing environment which manifests itself at the molecular level in the inherent promiscuity present in many enzymes. Exploiting enzyme promiscuity to develop novel functionality has been the focus of much recent research effort, where directed evolution techniques can improve the catalytic performance of even very weakly active starting points to become commercially relevant enzymes. In this review, we discuss the biochemical role of enzymes in terms of specificity and functionality, and then focus on recent developments in the application of structural bioinformatics methods to understand the evolution of specificity and guide *de novo* enzyme design.

Functional changes in enzyme evolution

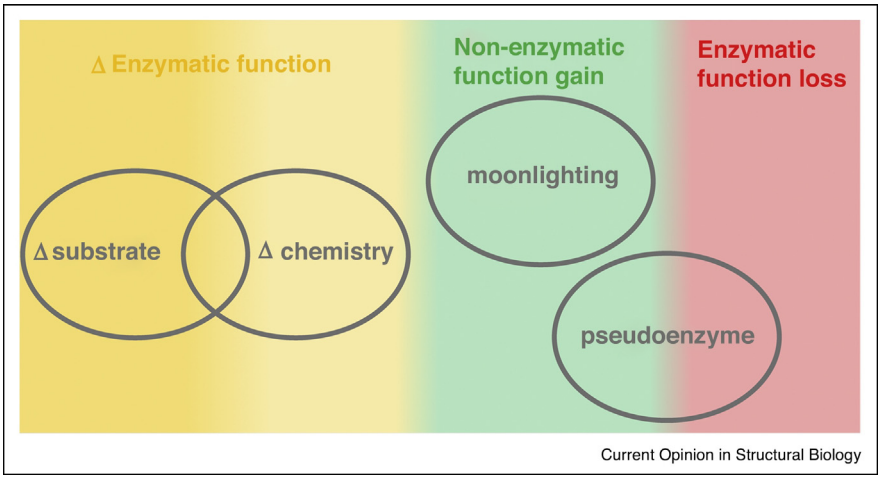
Evolution is a random generator of possible improvements in the face of the environmental challenges an organism experiences, where survival of the fittest ensures the retention of successful solutions into future generations. Evolution is powerful and has produced enzymes with varying degrees of substrate and function specificity where beneficial to the organism. Figure 1 shows the various types of functional changes observed in enzyme evolution, where more common changes in chemistry and substrates are supplemented with rarer gain or loss of function in the form of moonlighting and pseudo-enzymes respectively. These rarer functional changes will be considered briefly first, before the discussion moves to the more common evolutionary pathways.

Gain and loss of enzyme function: moonlighting and pseudo-enzymes

The acquisition of new functionality within an enzyme family as a result of gene duplication and mutation is commonly observed, see [1•] for some selected examples, but some enzymes can exhibit secondary, markedly different non-enzymatic functionality, when an enzyme with exactly the same chemical structure can moonlight to perform different roles in different cell compartments or environments [2,3]. These moonlighting enzymes, although rare, are being increasingly documented, where secondary function may be controlled by the varying ligand concentrations, by local phosphorylation levels and their different homo/hetero oligomeric states.

The additional functionality from moonlighting enzymes usually arises from a different site in the same structure and is distinct from gene fusions, multiple RNA splice variants or pleiotropic effects (where one gene influences two or more seemingly unrelated phenotypic traits), which can all affect enzyme catalysis. The evolution of secondary functional sites on an enzyme and the

Figure 1



Different types of functional changes in enzymes.

regulation of expression are active research questions [4] and demonstrate that multi-functional proteins are a design possibility, opening up opportunities for multi-functional polypeptide drugs and synthetic pathways. Researchers are becoming increasingly aware of the contribution of moonlighting enzymes but it is challenging to identify moonlighting functionality with bioinformatics methods [5] where small changes in structure or context can have dramatic effects on functionality [6]. Most discoveries of moonlighting occur from observation and serendipity and a database of moonlighting enzymes with approaching 300 entries has recently been curated from the literature [7], where one example is alpha-crystallin, the structural protein in the lens of the eye, that also has lactate dehydrogenase and argininosuccinate lyase activity.

Pseudo-enzymes are proteins that closely resemble an active enzyme, but at some point have lost their catalytic functionality and are retained in a genome for the beneficial new functionality that they have acquired [8], such as roles in regulatory and signaling pathways. Pseudo-enzymes are considered in more detail elsewhere in this edition, and the discussion here will move on to on the evolution of different functionality from the same binding pocket giving changes in substrate specificity and chemistry.

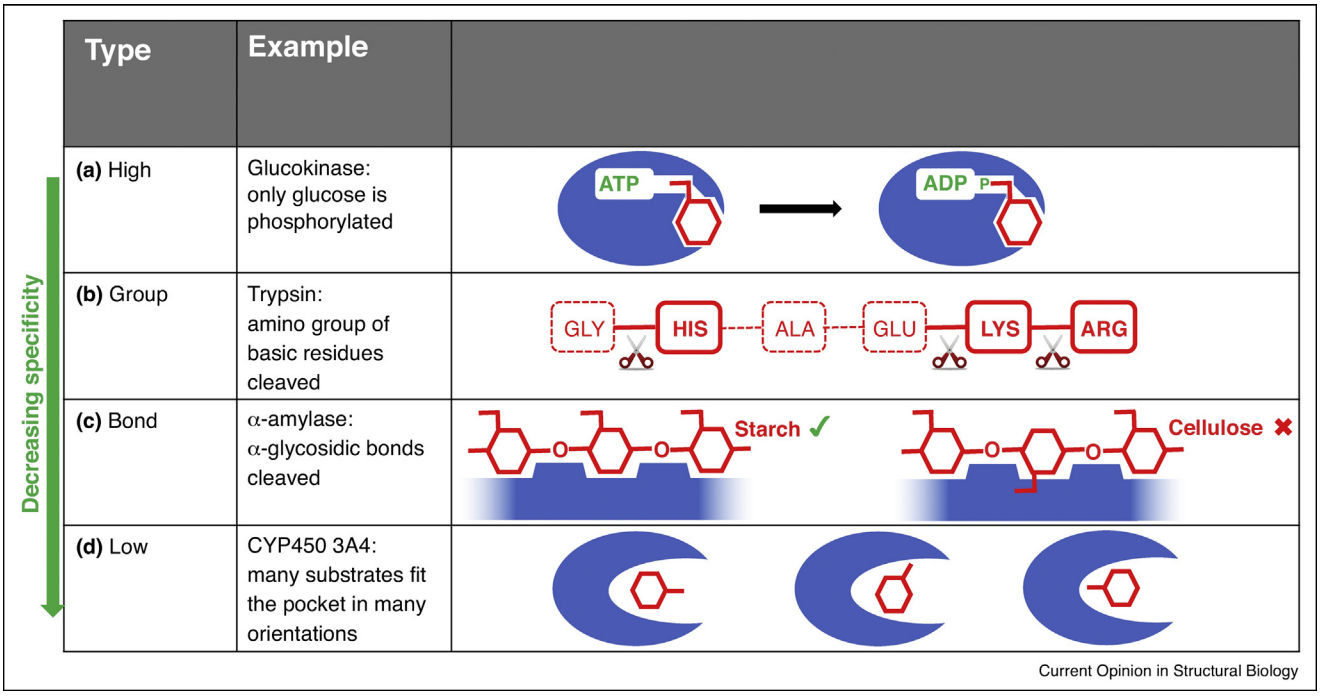
Different types of enzyme substrate specificity

Enzymes are able to give many orders of magnitude speed-up in essential reactions such as respiration, digestion and photosynthesis by stabilising transition states, thereby reducing activation energies and enabling reactions to proceed on timescales that can support life [9]. At a fundamental level, catalytic residues must be positioned around substrates in the correct orientations to stabilise transition states [10]. However, the specificity of the binding event between enzyme and substrate can vary depending on the extent to which the pocket has been optimised over evolutionary time [11]. An enzyme only needs to offer selectivity over detrimental side-reactions on potential substrates it is likely to encounter in its expressed location, where it becomes beneficial for evolution to deliver greater specificity. Indeed, some enzymes have evolved group or bond specificity such as those acting on some proteins and carbohydrates (see Figure 2 for trypsin and amylase examples), a more efficient solution then having to evolve a set of highly specific enzymes for every occurrence of each bond or group.

Enzyme specificity, defined according to the range of substrates and their similarity either for an individual enzyme or family of enzymes, exists on a continuum between highly specific and highly unspecific (promiscuous), demonstrating the concept that enzymes only evolve specificity when it is advantageous for the organism. It is challenging to define mutually exclusive categories to characterise the varied specificity observed, but despite this four categories of enzyme specificity have emerged [12]:

a) high specificity (an enzyme catalyzes one reaction at one site in one substrate to produce one product)
b) group specificity (an enzyme acts on a specific group (i.e. a given bond (cleaving or ligating) in a defined and restricted molecular environment))
c) bond specificity (an enzyme acts on a specific bond regardless of molecular environment)
d) low specificity (an enzyme can act at multiple sites in multiple substrates where site of reaction is influenced but not dictated by reactivity and accessibility considerations).

Figure 2



Categories of enzyme specificity with examples.

A classic example of low specificity is isoform 3A4 from the Cytochrome P450 family of enzymes that is involved in the oxidation of a broad range of xenobiotics at multiple sites in the substrate via several pathways, facilitating their subsequent conjugation and elimination from the organism [13]. Across the P450 family, the different isoforms have varying expression levels, tissue distributions, binding pocket sizes and characteristics giving rise to markedly different substrate profiles, causing them to occupy different positions on the specificity continuum.

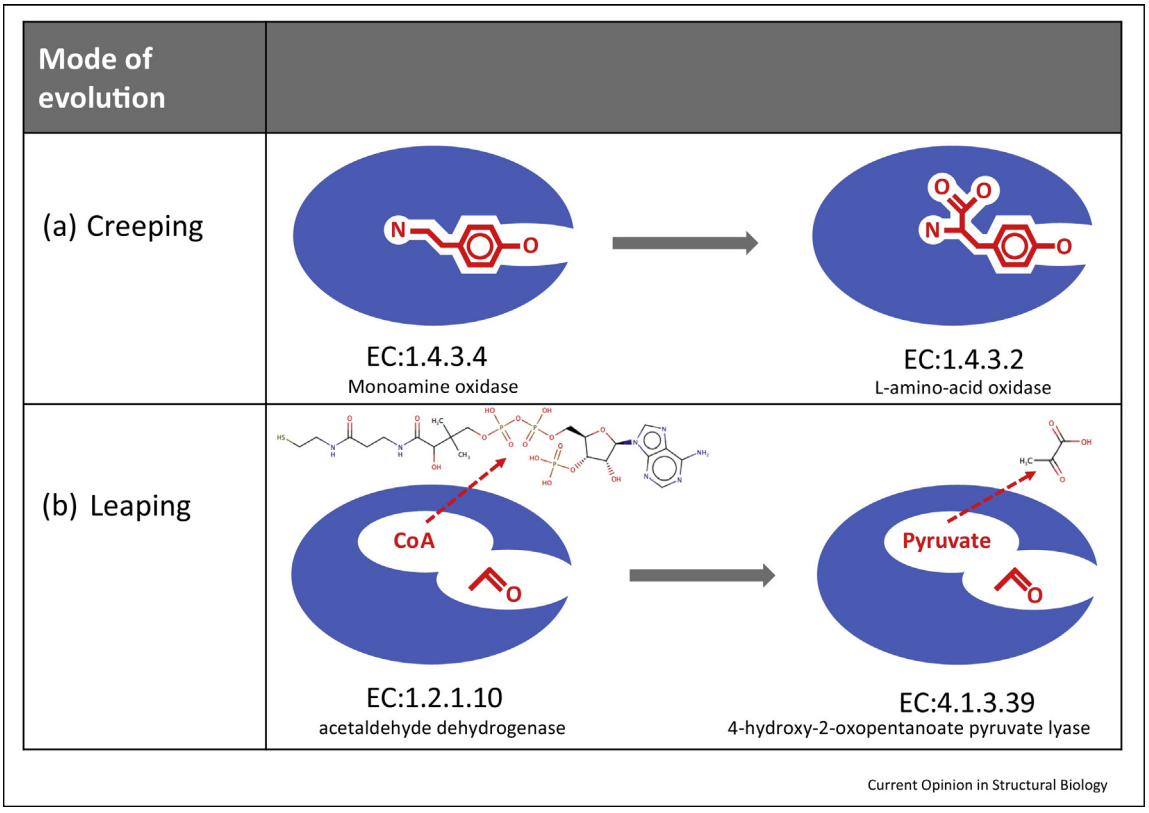
There is a further important consideration given the polymeric nature of the molecules of life. Many enzymes which act on proteins, DNA, RNA, sugar chains and lipids, show bond and group specificity, but polymer promiscuity, that is, although they have some side-chain specificity, they act on multiple different polymers. It would be extremely inefficient to have separate enzymes operating on all possible monomer combinations. For example, the alpha amylase cleavage of glycosidic links in starch and glycogen is bond specific [14], but interestingly also stereo specific and unable to cleave beta links in cellulose, disqualifying cellulose as a source of glucose for animals. Examples of group specific enzymes include pepsin, trypsin and chymotrypsin cleaving the amino groups of aromatic, amino groups of basic and carboxyl groups of aromatic amino acids respectively [15]. The phosphorylation of hexoses by hexokinase is also group specific, in contrast to the highly specific glucokinase acting only on glucose [16].

Modes of evolution: creeping and leaping

Enzyme evolution is complex [17••] where evolutionary events such as single-point mutations, indels and domain fusions can cause substrate specificity creep due to changes to the binding cavity or more remarkable leaps in chemistry often due to changes to catalytic residues. Most evolutionary changes cause a relatively minor change in function described as 'creeping evolution' but occasionally there can be a radical shift or 'leap' in function, for example, when a change allows the binding of a completely different substrate or when a substrate binds in 'reverse mode' or when the change provides an alteration in enzyme mechanism (Figure 3).

The core structure of the protein is rarely changed by such evolutionary events despite low sequence identity between relatives, providing a robust scaffold that enables changes in different relatives to give varying degrees of diverse chemistry [18••]. The focus of the remainder of this review is on the application of bioinformatics methods to understand enzyme promiscuity and the evolution of substrate specificity for a given catalytic site, usually confined to a single domain of the protein. However, nature likes to recycle and the emergence of novel functionality from the recombination of different

Figure 3



The creeping and leaping modes of evolution. Creeping involves incremental changes to the binding cavity leading to minor changes in functionality (example based on binding pocket mutations to give the substrate creep from EC:1.4.3.4 (monoamine oxidase) to EC:1.4.3.2 (L-amino-acid oxidase)). Leaping involves a radical shift in function, dramatically altering substrate binding or chemistry (example based on binding pocket mutations to give the leap between EC:1.2.1.10 (acetaldehyde dehydrogenase) and EC:4.1.3.39 (4-hydroxy-2-oxopentanoate pyruvate lyase)). Inset molecular structures generated with MarvinSketch [35].

domains [19,20] giving rise to new protein–protein interactions and quaternary structure [21,22] is also commonly observed, but not discussed further here.

The large scale study of enzyme evolution is becoming possible from the increasing amount of sequence and structural data available [1•] where alignment methods can be applied to identify evolutionary relationships between enzymes [23•]. It has been observed that changes in substrate specificity, probably arising from incremental binding site mutations, are far more likely than changes in chemistry, which probably require many complementary mutations to key catalytic residues without disrupting enzyme activity. However, the impact of longer range mutations further from the catalytic centre can sometimes have dramatic effects on the binding site meaning that the subtle but cumulative effects of second and third shell mutations cannot be ignored [24,25]. There is a growing body of evidence that non-additive interactions amongst mutations (epistasis) are common in adaptive evolution [26–28], amplifying the effect of later mutations and making the fitness landscape more extreme, enabling more diverse evolutionary trajectories to be accessed [29].

However, whilst substrate changes are more common, the capability of evolution to deliver dramatic leaps in chemistry is demonstrated by the remarkable observation that all changes between EC (Enzyme Commission) classes are observed. Most EC classes retain the same primary class, but isomerases EC5 are exceptional in that they are more likely to evolve a different function than remain an isomerase [30], with conversions to lyases EC4 occurring more frequently than expected [31].

The size of the evolutionary steps required to deliver these evolutionary creeps and leaps can be measured from the extent of the change in sequence and structure [30]. In order to explore the relationship between sequence changes and functional changes, it is necessary to develop quantitative measurements of changes in function and changes in specificity. The change in chemistry can be measured using recently developed computational methods to compare the similarity of enzyme reactions by

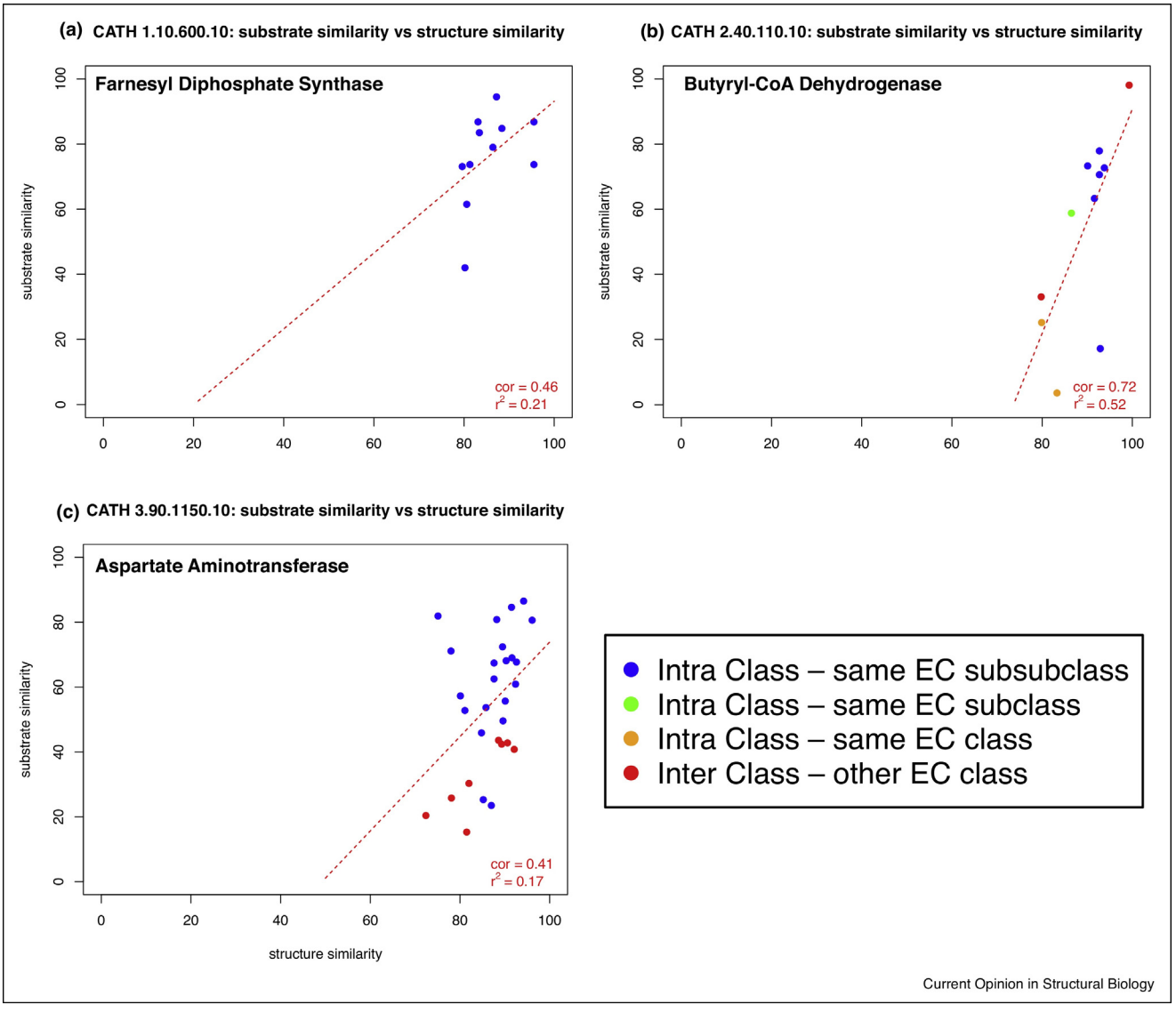
decomposing them into feature vectors derived from the bond changes they catalyze, and the reaction centres and substrates they operate on [32•,33•]. This allows the correlation between change in substrate structure and change in enzyme structure/sequence to be investigated.

The ability to quantitatively measure enzyme reaction similarity also reveals inconsistencies and irregularities inherent in the current hierarchical EC numbering system. For instance, almost a third of all known EC numbers are associated with more than one enzyme reaction in the KEGG database [34].

Measuring changes in specificity and function

Recent work in our group has used FunTree [23•] and CATH [36] to identify evolutionary relationships between enzymes within a homologous family and investigate the correlation between change in substrate structure (calculated using ECBLAST [32•]) and change in enzyme structure (calculated using SSAP [37•]) generating plots at the CATH domain level. An example domain has been chosen from each of the CATH structural classes (a) All Alpha, (b) All Beta and (c) Mixed Alpha/Beta, with Figure 4 showing the analysis for CATH domains (a) 1.10.600.10 (e.g. farnesyl diphosphate

Figure 4



Substrate similarity of the main reactant (calculated using ECBLAST) against domain structure similarity (calculated using SSAP) for each evolutionary relationship identified with FunTree for CATH domains: **(a)** 1.10.600.10, **(b)** 2.40.110.10, **(c)** 3.90.1150.10. One member enzyme of each domain family is shown in the top left.

synthase), (b) 2.40.110.10 (e.g. butyryl-CoA dehydrogenase) and (c) 3.90.1150.10 (e.g. aspartate aminotransferase), where each data point is color coded depending on the EC level at which the change occurs. For each example domain we have fitted a line to the points to gain an approximate correlation.

Inspecting the plots for all domains it is clear that a trend exists whereby bigger changes in substrate structure generally require bigger changes in protein sequence and structure, although there are some exceptions where the linear trend is broken and the data points form a cluster. Also, some domains apparently show an anti-correlation, where the most similar substrates are catalyzed by the most dissimilar enzymes. Each family is complex and requires detailed analysis. Examining the evolution within each CATH domain allows comparisons to be made in terms of the number of evolutionary changes observed, the diversity of the chemistry performed, the number of changes in function (as defined by a change in EC class) and the amount of structural change required to accommodate new substrates and mechanisms. This will help to give insights into the evolvability of different domains and may help in understanding which domains have the potential to be multi-specific in terms of finding suitable starting points for directed evolution.

For instance, in the examples shown, CATH domain 1.10.600.10 only shows changes in substrate, whereas domain 2.40.110.10 shows changes at all EC levels, including changes in the bonds broken/formed, the basic reaction and the substrate specificity and domain 3.90.1150.10 shows changes of substrate (4th level EC) and changes at the primary EC class level. Domain 3.90.1150.10 contains 29 data points, almost three times as many as the other domains, which may indicate a more versatile and evolvable domain.

What determines specificity?

Understanding enzyme evolution and the factors that facilitate specificity would be helpful in *de novo* enzyme design and it has been proposed that a key enabler of promiscuity is the flexibility of the enzyme [38]. Evolution gives cycles of destabilisation and restabilisation allowing conformational space to be sampled [39] whilst maintaining the positions of important catalytic residues. The trade-off between stability and activity is demonstrated in the adaptive evolution of RubisCO [40] and highlights the strong biophysical constraints that influence the evolution of enzymes.

It has also been suggested that the location of active site residues can influence the evolvability. The presence of active site residues in flexible, loosely packed loops distinct from the core scaffold is thought to facilitate high evolvability [41••] where mutations to key residues are less likely to have a destabilising effect on the protein. For example, in TIM barrels, the location of the active site at one end of the barrel involving many loops allows them to incorporate more easily a wide range of cofactors and contributes to their prevalence in modern proteomes where they support very diverse substrates and chemistry [42].

A further important feature to facilitate promiscuous function may be the presence of water networks to stabilise binding of non-native substrates, which, if providing advantageous functionality, would be refined over evolutionary time by the replacement of entropically unfavourable water-mediated interactions with polar protein–ligand interactions [43].

Resurrecting ancestral enzymes

Understanding natural evolution gives opportunities to resurrect ancestral enzymes as suitable starting points for directed evolution [44–46]. It has been proposed that early enzymes were generalists with broad functionality and substrate specificity [47••], giving rise to more specialist enzymes over evolutionary time. However, it is also possible that enzymes could evolve to be more promiscuous given the right circumstances. The stability of early enzymes in the face of harsher prevailing conditions [48] might also make them suitable candidates for repurposing and more tolerant to the high mutational load of directed evolution.

The evolution of specificity from a common ancestor is demonstrated by the flavin-dependent monooxygenases where a promiscuous FAD binding domain gave rise to more specific functionality from various domain fusion events [49]. Domain fusion events are also implicated in the structurally and functionally diverse HADSF enzyme superfamily where enzymes with an inserted CAP domain show wider substrate promiscuity [50]. The Cytochrome P450s also have a broad substrate range making them repurposing candidates for drugs and pharmaceuticals [51].

Exploiting enzyme promiscuity

The innate substrate promiscuity of many enzymes, showing weak activity with off-target but similar molecules, can help to drive the acquisition of new functions. This inherent promiscuity exists since over evolutionary time, the pressure of natural selection ceases when further catalytic or specificity improvements do not improve fitness [52••]. Therefore, a perfectly specific active site is unnecessary and in a changing environment it is likely to be beneficial for enzymes to retain an inherent promiscuity and the corresponding ability to evolve new functions. This is demonstrated by the acquisition of resistance to beta-lactam antibiotics by populations of bacteria over many generations, one of the key challenges to modern medicine, where progress has been made in understanding the resistance profile of the different

beta-lactamases and pinpointing resistance properties to key sequence sites [53].

Promiscuity is thought to be one factor that facilitates natural evolution, and can be exploited by directed evolution [54,55] which has brought enzyme catalyzed synthesis to many industries including pharmaceuticals, textiles and food [56–58]. Promiscuity can vary greatly between orthologous enzymes [59] as a result of neutral sequence divergence so it is important to consider multiple orthologues as start points in directed evolution.

The importance and diversity of directed evolution is demonstrated by recent progress in developing enzymes to improve CO2 fixation [45,60]; detoxify organophosphates from contaminated soil and water [61]; catalyze the formation of organosilicon compounds [62]; and destroy latent HIV pro-virus in cells [63]. Directed evolution methods reinforce the fact that there is nothing magical about the honing of mechanisms by natural evolution, and sophisticated, highly active and novel mechanisms have been shown to emerge from ultra-high-throughput techniques, such as the emergence of a catalytic tetrad to rival the efficiencies of natural aldolases [64].

Opportunities for the future

Enzyme promiscuity is an important factor in enzyme evolution where new functions emerge at the edges of current functionality from the refinement of weak non-native interactions as required by environmental demands. The development of *de novo* enzymes using directed evolution aims to exploit this inherent promiscuity and bioinformatics methods can be used to identify suitable start points and guide these approaches. Any insights that can be garnered from natural evolution to inform *de novo* enzyme design are invaluable, such as the identification of evolutionary labile structures and molecules that are likely to show some promiscuous activity with current enzymes.

Quantitative measures of substrate specificity and promiscuity would be useful additions to the meta data associated with PDB structures, helping to guide biologists in their selection of starting points for *de novo* enzyme design. Understanding the evolution of function and molecular mechanism may also help to generate more knowledge for enzyme design.

The study of the evolution of substrate specificity is an active avenue of research attempting to link the flexibility, modularity, and stability of enzyme structure to function and fitness. The potential of exploiting enzyme promiscuity to give commercial catalysts of outstanding efficiency and specificity is beginning to be fully realized.

Conflict of interest

Nothing declared.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. • Furnham N, Dawson NL, Rahman SA, Thornton JM, Orengo CA: **Large-scale analysis exploring evolution of catalytic machineries and mechanisms in enzyme superfamilies**. *J Mol Biol [Internet]* 2016, **428**:253-267 Available from: http://linkinghub.elsevier.com/retrieve/pii/S0022283615006531.
One of the first enzyme evolution studies to apply bioinformatics methods to analyze change in function and catalytic machinery across structural domains on a large-scale basis.
2. Jeffery CJ: **Why study moonlighting proteins?** *Front Genet [Internet]* 2015:6. Available from: http://journal.frontiersin.org/Article/10.3389/fgene.2015.00211/Abstract.
3. Henderson B: *Moonlighting Proteins: Novel Virulence Factors in Bacterial Infections*. Wiley-Blackwell; 2017. p. 472.
4. Espinosa-Cantú A, Ascencio D, Barona-Gómez F, DeLuna A: **Gene duplication and the evolution of moonlighting proteins**. *Front Genet [Internet]* 2015:6. Available from: http://journal.frontiersin.org/Article/10.3389/fgene.2015.00227/Abstract.
5. Hernández S, Franco L, Calvo A, Ferragut G, Hermoso A, Amela I *et al.*: **Bioinformatics and moonlighting proteins**. *Front Bioeng Biotechnol [Internet]* 2015:3. Available from: http://journal.frontiersin.org/Article/10.3389/fbioe.2015.00090/Abstract.
6. Jeffery CJ: **Protein species and moonlighting proteins: Very small changes in a protein's covalent structure can change its biochemical function**. *J Proteomics [Internet]* 2016, **134**:19-24 Available from: http://linkinghub.elsevier.com/retrieve/pii/S1874391915301469.
7. Mani M, Chen C, Amblee V, Liu H, Mathur T, Zwicke G *et al.*: **MoonProt: a database for proteins that are known to moonlight**. *Nucleic Acids Res [Internet]* 2015, **43** D277-82. Available from: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gku954.
8. Eyers PA, Murphy JM: **The evolving world of pseudoenzymes: proteins, prejudice and zombies**. *BMC Biol [Internet]* 2016, **14**:98 Available from: http://bmcbiol.biomedcentral.com/articles/10.1186/s12915-016-0322-x.
9. Bugg T: *Introduction to Enzyme and Coenzyme Chemistry*. edn 3. Wiley; 2012.
10. Bartlett GJ, Porter CT, Borkakoti N, Thornton JM: **Analysis of catalytic residues in enzyme active sites**. *J Mol Biol [Internet]*. 2002, **324**:105-121 Available from: http://linkinghub.elsevier.com/retrieve/pii/S0022283602010367.
11. Nobeli I, Favia AD, Thornton JM: **Protein promiscuity and its implications for biotechnology**. *Nat Biotechnol [Internet]* 2009, **27**:157-167 Available from: http://www.nature.com/doifinder/10.1038/nbt1519.
12. Freehold NJ: *Manual of Clinical Enzyme Measurements*. Worthingt Biomed Corp.; 1972.
13. Kirchmair J, Williamson MJ, Tyzack JD, Tan L, Bond PJ, Bender A *et al.*: **Computational prediction of metabolism: sites, products, SAR, P450 enzyme dynamics, and mechanisms**. *J Chem Inf Model [Internet]* 2012, **52**:617-648 Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3317594&tool=pmcentrez&rendertype=Abstract.
14. Janeček , Svensson B, MacGregor EA: **α-Amylase: an enzyme specificity found in various families of glycoside hydrolases**. *Cell Mol Life Sci [Internet]* 2014, **71**:1149-1170 Available from: http://link.springer.com/10.1007/s00018-013-1388-z.
15. Keil B: *Specificity of Proteolysis*. Springer Science & Business Media; 2012.

16. Richter JP, Goroncy AK, Ronimus RS, Sutherland-Smith AJ: **The structural and functional characterization of mammalian ADP-dependent glucokinase**. *J Biol Chem [Internet]* 2016, **291**:3694-3704 Available from: http://www.jbc.org/lookup/doi/10.1074/jbc.M115.679902.

17. •• Baier F, Copp JN, Tokuriki N: **Evolution of enzyme superfamilies: comprehensive exploration of sequence–function relationships**. *Biochemistry [Internet]* 2016, **55**:6375-6388 Available from: http://pubs.acs.org/doi/abs/10.1021/acs.biochem.6b00723.
A review of integrative approaches to explore enzyme evolution and functional diversity within enzyme superfamilies, highlighting recent large-scale experimental methods to determine activity profiles across enzyme superfamilies.

18. •• Brown SD, Babbitt PC: **New insights about enzyme evolution from large scale studies of sequence and structure relationships**. *J Biol Chem [Internet]* 2014, **289**:30221-30228 Available from: http://www.jbc.org/lookup/doi/10.1074/jbc.R114.569350.
Highlights the power of using similarity networks to give insights about enzyme evolution from sequence and structural data.

19. Lees JG, Dawson NL, Sillitoe I, Orengo CA: **Functional innovation from changes in protein domains and their combinations**. *Curr Opin Struct Biol [Internet]* 2016, **38**:44-52 Available from: http://linkinghub.elsevier.com/retrieve/pii/S0959440X16300513.

20. Hsu C-H, Chiang AWT, Hwang M-J, Liao B-Y: **Proteins with highly evolvable domain architectures are nonessential but highly retained**. *Mol Biol Evol [Internet]* 2016, **33**:1219-1230 Available from: http://mbe.oxfordjournals.org/lookup/doi/10.1093/molbev/msw006.

21. Marsh JA, Teichmann SA: **Structure, dynamics, assembly, and evolution of protein complexes**. *Annu Rev Biochem [Internet]* 2015, **84**:551-575 Available from: http://www.annualreviews.org/doi/10.1146/annurev-biochem-060614-034142.

22. Laskowski RA, Thornton JM: **Proteins: interaction at a distance**. *IUCrJ [Internet]* 2015, **2**:609-610 Available from: http://scripts.iucr.org/cgi-bin/paper?S2052252515020217.

23. • Sillitoe I, Furnham N: **FunTree: advances in a resource for exploring and contextualising protein function evolution**. *Nucleic Acids Res [Internet]* 2016, **44**:D317-D323 Available from: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkv1274.
Funtree is a unique resource bringing together protein sequence, structure and function information that allows the investigation of novel functionality within a structurally defined superfamily.

24. Yang G, Hong N, Baier F, Jackson CJ, Tokuriki N: **Conformational tinkering drives evolution of a promiscuous activity through indirect mutational effects**. *Biochemistry [Internet]* 2016, **55**:4583-4593 Available from: http://pubs.acs.org/doi/abs/10.1021/acs.biochem.6b00561.

25. Jack BR, Meyer AG, Echave J, Wilke CO: **Functional sites induce long-range evolutionary constraints in enzymes**. *PLoS Biol [Internet]* 2016, **14**:e1002452 Available from: http://dx.plos.org/10.1371/journal.pbio.1002452.

26. Miton CM, Tokuriki N: **How mutational epistasis impairs predictability in protein evolution and design**. *Protein Sci [Internet]* 2016, **25**:1260-1272 Available from: http://doi.wiley.com/10.1002/pro.2876.

27. Starr TN, Thornton JW: **Epistasis in protein evolution**. *Protein Sci [Internet]* 2016, **25**:1204-1218 Available from: http://doi.wiley.com/10.1002/pro.2897.

28. Cheema J, Faraldos JA, O'Maille PE: **Review: Epistasis and dominance in the emergence of catalytic function as exemplified by the evolution of plant terpene synthases**. *Plant Sci [Internet]* 2017, **255**:29-38 Available from: http://linkinghub.elsevier.com/retrieve/pii/S0168945216306495.

29. Steinberg B, Ostermeier M: **Shifting fitness and epistatic landscapes reflect trade-offs along an evolutionary pathway**. *J Mol Biol [Internet]* 2016, **428**:2730-2743 Available from: http://linkinghub.elsevier.com/retrieve/pii/S0022283616301450.

30. Martínez Cuesta S, Rahman SA, Furnham N, Thornton JM: **The classification and evolution of enzyme function**. *Biophys J [Internet]* 2015, **109**:1082-1086 Available from: http://linkinghub.elsevier.com/retrieve/pii/S0006349515004002.

31. Martínez Cuesta S, Rahman SA, Thornton JM: **Exploring the chemistry and evolution of the isomerases**. *Proc Natl Acad Sci [Internet]* 2016, **113**:1796-1801 Available from: http://www.pnas.org/lookup/doi/10.1073/pnas.1509494113.

32. • Rahman SA, Cuesta SM, Furnham N, Holliday GL, Thornton JM: **EC-BLAST: a tool to automatically search and compare enzyme reactions**. *Nat Methods [Internet]* 2014, **11**:171-174 Available from: http://www.nature.com/doifinder/10.1038/nmeth.2803.
ECBLAST is a web tool to perform quantitative similarity comparisons between enzyme reactions at three levels: bond change, reaction center and compound structure.

33. • Rahman SA, Torrance G, Baldacci L, Martínez Cuesta S, Fenninger F, Gopal N *et al.*: **Reaction Decoder Tool (RDT): extracting features from chemical reactions**. *Bioinformatics [Internet]* 2016, **32**:2065-2066 Available from: https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btw096.
Reaction Decoder Tool is a command-line tool to quantify the chemical similarity between reactions at three levels: bond change, reaction center and compound structure. It can be applied to quantify the size of evolutionary steps from a functional perspective.

34. Dönertaş HM, Martínez Cuesta S, Rahman SA, Thornton JM: **Characterising complex enzyme reaction data**. *PLoS One [Internet]* 2016, **11**:e0147952 Available from: http://dx.plos.org/10.1371/journal.pone.0147952.

35. ChemAxon. MarvinSketch, Version 16.6.13.0. 2016.

36. Sillitoe I, Lewis TE, Cuff A, Das S, Ashford P, Dawson NL *et al.*: **CATH: comprehensive structural and functional annotations for genome sequences**. *Nucleic Acids Res [Internet]* 2015, **43**:D376-D381 Available from: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gku947.

37. • Orengo CA, Taylor WR: **A local alignment method for protein structure motifs**. *J Mol Biol [Internet]* 1993, **233**:488-497 Available from: http://linkinghub.elsevier.com/retrieve/pii/S0022283683715263.
This software performs and scores structural alignments. It can be applied to quantify the size of evolution steps from a structural perspective.

38. Campbell E, Kaltenbach M, Correy GJ, Carr PD, Porebski BT, Livingstone EK *et al.*: **The role of protein dynamics in the evolution of new enzyme function**. *Nat Chem Biol [Internet]* 2016, **12**:944-950 Available from: http://www.nature.com/doifinder/10.1038/nchembio.2175.

39. Ma B, Nussinov R: **Protein dynamics: conformational footprints**. *Nat Chem Biol [Internet]* 2016, **12**:890-891 Available from: http://www.nature.com/doifinder/10.1038/nchembio.2212.

40. Studer RA, Christin P-A, Williams MA, Orengo CA: **Stability–activity tradeoffs constrain the adaptive evolution of RubisCO**. *Proc Natl Acad Sci [Internet]* 2014, **111**:2223-2228 Available from: http://www.pnas.org/lookup/doi/10.1073/pnas.1310811111.

41. •• Dellus-Gur E, Toth-Petroczy A, Elias M, Tawfik DS: **What makes a protein fold amenable to functional innovation? Fold polarity and stability trade-offs**. *J Mol Biol [Internet]* 2013, **425**:2609-2621 Available from: http://linkinghub.elsevier.com/retrieve/pii/S0022283613002003.
Proposes a rationalization for enzyme fold evolvability in terms of the connectivity of the core scaffold to the active site making some structures inherently more predisposed to the evolution of multiple functions.

42. Goldman AD, Beatty JT, Landweber LF, The TIM: **Barrel architecture facilitated the early evolution of protein-mediated metabolism**. *J Mol Evol [Internet]* 2016, **82**:17-26 Available from: http://link.springer.com/10.1007/s00239-015-9722-8.

43. Clifton BE, Jackson CJ: **Ancestral protein reconstruction yields insights into adaptive evolution of binding specificity in solute-binding proteins**. *Cell Chem Biol [Internet]* 2016, **23**:236-245 Available from: http://linkinghub.elsevier.com/retrieve/pii/S2451945616000313.

44. Alcalde M: **When directed evolution met ancestral enzyme resurrection**. *Microb Biotechnol [Internet]* 2016. Available from: http://doi.wiley.com/10.1111/1751-7915.12452.

45. Shih PM, Occhialini A, Cameron JC, Andralojc PJ, Parry MAJ, Kerfeld CA: **Biochemical characterization of predicted Precambrian RuBisCO**. *Nat Commun [Internet]* 2016, **7**:10382 Available from: http://www.nature.com/doifinder/10.1038/ncomms10382.

46. Alcalde M: **Engineering the ligninolytic enzyme consortium**. *Trends Biotechnol [Internet]* 2015, **33**:155-162 Available from: http://linkinghub.elsevier.com/retrieve/pii/S016777991400256X.

47. •• Khersonsky O, Tawfik DS: **Enzyme promiscuity: a mechanistic and evolutionary perspective**. *Annu Rev Biochem [Internet]* 2010, **79**:471-505 Available from: http://www.annualreviews.org/doi/10.1146/annurev-biochem-030409-143718.

Discusses the structural, mechanistic and evolutionary implications of promiscuity including the different ways enzymes exhibit promiscuous interactions.

48. Wheeler LC, Lim SA, Marqusee S, Harms MJ: **The thermostability and specificity of ancient proteins**. *Curr Opin Struct Biol [Internet]* 2016, **38**:37-43 Available from: http://linkinghub.elsevier.com/retrieve/pii/S0959440X16300501.

49. Mascotti ML, Juri Ayub M, Furnham N, Thornton JM, Laskowski RA: **Chopping and changing: the evolution of the flavin-dependent monooxygenases**. *J Mol Biol [Internet]* 2016, **428**:3131-3146 Available from: http://linkinghub.elsevier.com/retrieve/pii/S0022283616302480.

50. Huang H, Pandya C, Liu C, Al-Obaidi NF, Wang M, Zheng L *et al.*: **Panoramic view of a superfamily of phosphatases through substrate profiling**. *Proc Natl Acad Sci [Internet]* 2015, **112**:E1974-E1983 Available from: http://www.pnas.org/lookup/doi/10.1073/pnas.1423570112.

51. Behrendorff JBYH, Huang W, Gillam EMJ: **Directed evolution of cytochrome P450 enzymes for biocatalysis: exploiting the catalytic versatility of enzymes with relaxed substrate specificity**. *Biochem J [Internet]* 2015, **467**:1-15 Available from: http://biochemj.org/lookup/doi/10.1042/BJ20141493.

52. •• Copley SD: **An evolutionary biochemist's perspective on promiscuity**. *Trends Biochem Sci [Internet]* 2015, **40**:72-78 Available from: http://linkinghub.elsevier.com/retrieve/pii/S0968000414002217.

Defines and discusses enzyme promiscuity and the emergence of new functionality from promiscuous interactions.

53. Lee D, Das S, Dawson NL, Dobrijevic D, Ward J, Orengo C: **Novel computational protocols for functionally classifying and characterising serine beta-lactamases**. *PLoS Comput Biol [Internet]* 2016, **12**:e1004926 Available from: http://dx.plos.org/10.1371/journal.pcbi.1004926.

54. Tizei PAG, Csibra E, Torres L, Pinheiro VB: **Selection platforms for directed evolution in synthetic biology**. *Biochem Soc Trans [Internet]* 2016, **44**:1165-1175 Available from: http://biochemsoctrans.org/cgi/doi/10.1042/BST20160076.

55. Molina-Espeja P, Viña-Gonzalez J, Gomez-Fernandez BJ, Martin-Diaz J, Garcia-Ruiz E, Alcalde M: **Beyond the outer limits of nature by directed evolution**. *Biotechnol Adv [Internet]* 2016, **34**:754-767 Available from: http://linkinghub.elsevier.com/retrieve/pii/S0734975016300404.

56. Choi J-M, Han S-S, Kim H-S: **Industrial applications of enzyme biocatalysis: current status and future aspects**. *Biotechnol Adv [Internet]* 2015, **33**:1443-1454 Available from: http://linkinghub.elsevier.com/retrieve/pii/S0734975015000506.

57. Porter JL, Rusli RA, Ollis DL: **Directed evolution of enzymes for industrial biocatalysis**. *ChemBioChem [Internet]* 2016, **17**:197-203 Available from: http://doi.wiley.com/10.1002/cbic.201500280.

58. Otte KB, Hauer B: **Enzyme engineering in the context of novel pathways and products**. *Curr Opin Biotechnol [Internet]* 2015, **35**:16-22 Available from: http://linkinghub.elsevier.com/retrieve/pii/S0958166914002262.

59. Khanal A, Yu McLoughlin S, Kershner JP, Copley SD: **Differential effects of a mutation on the normal and promiscuous activities of orthologs: implications for natural and directed evolution**. *Mol Biol Evol [Internet]* 2015, **32**:100-108 Available from: http://mbe.oxfordjournals.org/cgi/doi/10.1093/molbev/msu271.

60. Erb TJ, Zarzycki J: **Biochemical and synthetic biology approaches to improve photosynthetic CO_2-fixation**. *Curr Opin Chem Biol [Internet]* 2016, **34**:72-79 Available from: http://linkinghub.elsevier.com/retrieve/pii/S1367593116300928.

61. Ng T-K, Gahan LR, Schenk G, Ollis DL: **Altering the substrate specificity of methyl parathion hydrolase with directed evolution**. *Arch Biochem Biophys [Internet]* 2015, **573**:59-68 Available from: http://linkinghub.elsevier.com/retrieve/pii/S000398611500140X.

62. Kan SBJ, Lewis RD, Chen K, Arnold FH: **Directed evolution of cytochrome c for carbon–silicon bond formation: bringing silicon to life**. *Science (80-) [Internet]* 2016, **354**:1048-1051 Available from: http://www.sciencemag.org/lookup/doi/10.1126/science.aah6219.

63. Karpinski J, Hauber I, Chemnitz J, Schäfer C, Paszkowski-Rogacz M, Chakraborty D *et al.*: **Directed evolution of a recombinase that excises the provirus of most HIV-1 primary isolates with high specificity**. *Nat Biotechnol [Internet]* 2016, **34**:401-409 Available from: http://www.nature.com/doifinder/10.1038/nbt.3467.

64. Obexer R, Godina A, Garrabou X, Mittl PRE, Baker D, Griffiths AD *et al.*: **Emergence of a catalytic tetrad during evolution of a highly active artificial aldolase**. *Nat Chem [Internet]* 2016. Available from: http://www.nature.com/doifinder/10.1038/nchem.2596.